MicroNet Challenge

Trevor Gale[1], Erich Elsen[2], Sara Hooker[1], Olivier Temam[2], Scott Gray[3], Jongsoo Park[4], Cliff Young[1], Utku Evci[1], Niki Parmar[1], Ashish Vaswani[1].
[1]Google, [2]DeepMind, [3]OpenAI, [4]Facebook

micronet-challenge.github.io

MicroNet Challenge

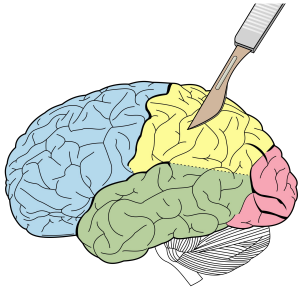
Large-scale model compression competition focused on efficient inference and future hardware. Contestants compete to design the most efficient models that solve the target task to the specified quality level.

Scoring

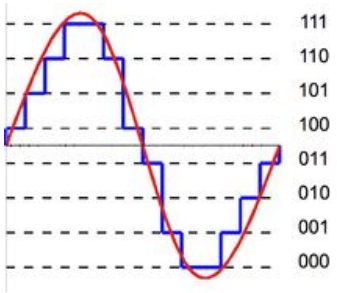
Our target metric is decoupled from existing hardware to incentivize exploration.

$$score = \frac{parameter\ storage}{baseline\ parameters} + \frac{math\ operations}{baseline\ ops}$$

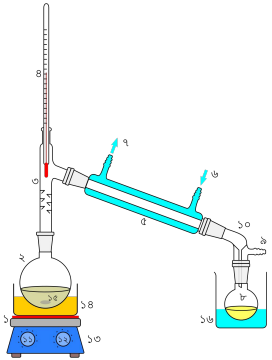
Parameter storage & math operations normalized by state-of-the-art baseline. Sparsity, quantization taken into account.



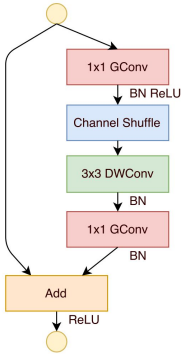
Weight Pruning



Quantization



Distillation

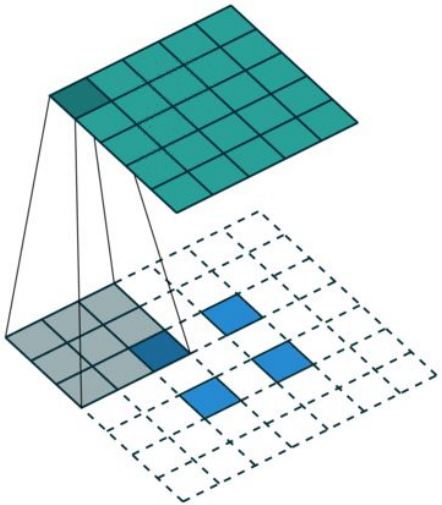


Building Blocks

Competition Tracks

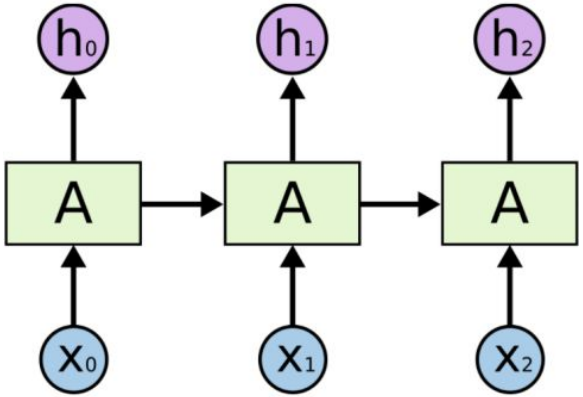
ImageNet

Image Classification
1.28M/50K train/dev images
75% top-1 accuracy



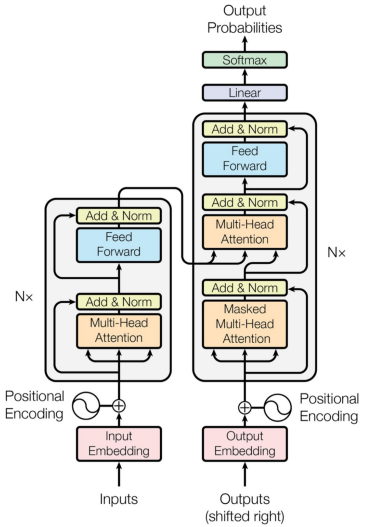
WikiText-103

Language Modeling
217K/245K train/dev words
35 test perplexity



CIFAR-100

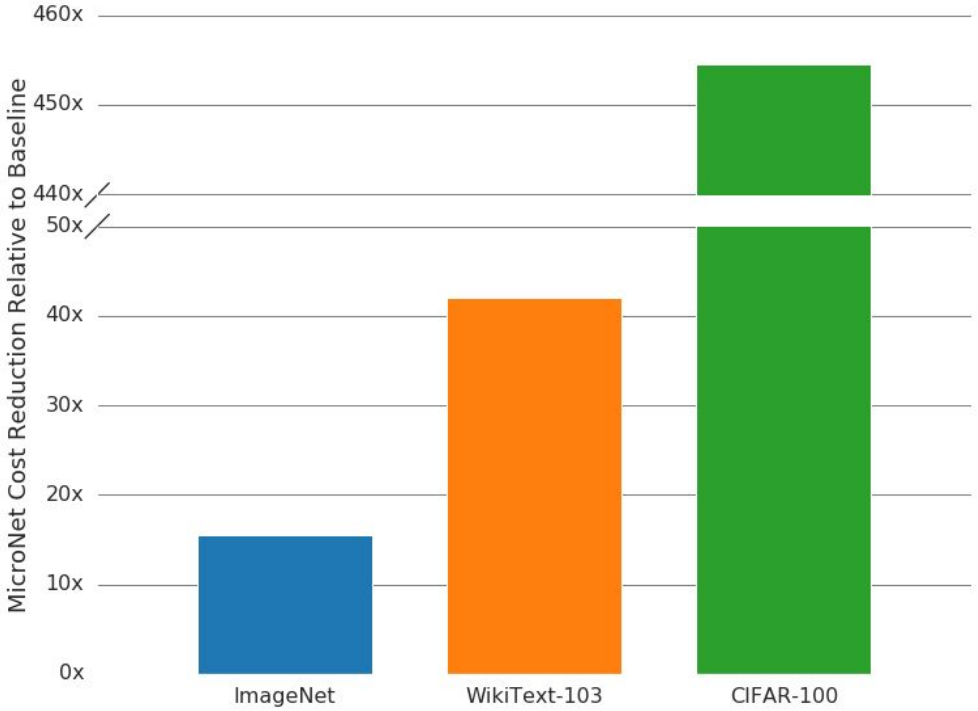
Image Classification
50K/10K train/dev images
80% top-1 accuracy



Results

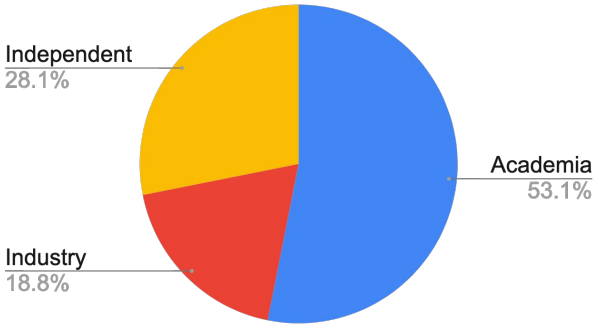
55 submissions from **32** unique organizations.

15x, **42x**, and **455x** improvements in our target metric over baseline[1] for ImageNet, WikiText-103, and CIFAR-100 models respectively.

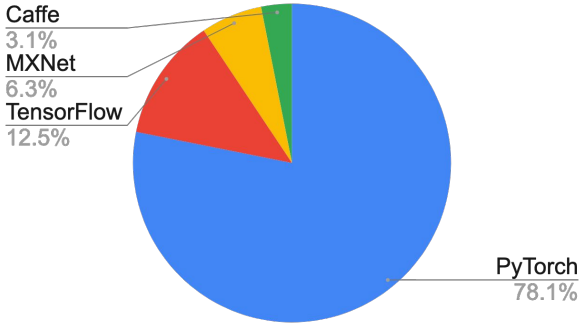


[1]baseline models are uncompressed, single-precision floating-point.

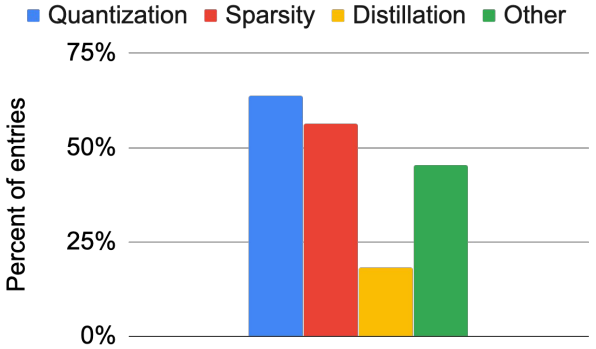
Submission Statistics



Affiliation Breakdown



Framework Breakdown



Technique Breakdown

Contestants represent a mix of different affiliations. PyTorch is the framework of choice by a wide margin. Wide variety of approaches for compression. Mixes of different compression approaches very popular.

Roadmap

An Empirical Study of Network Compression for Image Classification
Peisong Wang, Xiangyu He, Tianli Zhao, Cong Leng, Yifan Zhang and Jian Cheng
Winners of the ImageNet and CIFAR-100 tracks

Highlights From the Competition
Trevor Gale and Erich Elsen

Memory-Augmented Language Models with Network Compression
Zhongxia Yan and Hanrui Wang
Winners of the WikiText-103 track

Updates and Improvements for the 2020 MicroNet Challenge
Trevor Gale and Erich Elsen